BULLETIN DU MUSÉUM NATIONAL D'HISTOIRE NATURELLE
2e Série — Tome 40 — N° 3, 1968, pp. 465-473.

UTILISATION DES STATISTIQUES D'ORDRE EN TAXONOMIE NUMÉRIQUE

Par J. DAGET et J. C. HUREAU

Le but de la taxonomie numérique étant de faire intervenir le maximum de caractères distinctifs dans la recherche des similitudes entre unités taxonomiques opérationnelles, il est souvent nécessaire d'utiliser conjointement divers types de variables auxquelles toutes les opérations de l'arithmétique ne peuvent être appliquées indifféremment.

Sans entrer dans les détails, rappelons qu'il existe quatre types de variables :

— *nominales*, dont les différents états correspondent à des catégories qualitatives que l'on ne peut ranger en ordre croissant ou décroissant ;

— *ordinales*, dont les différents états peuvent être rangés de façon univoque dans un ordre croissant ou décroissant ;

— *repérables*, pour lesquelles il existe une échelle de référence dont l'unité et l'origine sont arbitraires ;

— *mesurables*, pour lesquelles il existe une échelle de mesure dont l'unité est arbitraire mais l'origine fixe.

Les différents états des variables repérables et mesurables s'expriment directement en valeurs numériques dont on peut calculer moyennes et variances. La standardisation, changement d'unité et d'origine destiné à rendre la moyenne nulle et la variance égale à l'unité, est donc praticable et le calcul des distances taxonomiques dans un hyperespace à n dimensions ou des intercorrélations par les coefficients de Bravais-Pearson est valable.

Par contre les différents états des variables ordinales et nominales doivent être codés et les valeurs codes choisies n'ont aucune signification arithmétique. Le calcul de la moyenne et de la variance effectué sur ces valeurs codes n'a aucun sens, les résultats étant fonction du système de codage arbitrairement choisi. Pour les variables nominales, seules des statistiques basées sur les critères de présence-absence ou d'association sont valables. Pour les variables ordinales, les *statistiques de rang* ou d'ordre sont également valables.

Dans la pratique, la Taxonomie numérique impose l'emploi d'un système de codage uniforme pour tous les caractères utilisés de façon à pouvoir les traiter simultanément comme un ensemble de variables ordinales. Ceci n'est possible pour les variable nominales que si elles ne présentent pas plus de deux états (présence-absence), mais on peut toujours s'arranger pour qu'il en soit ainsi. On notera dès maintenant que le fait de transformer des variables mesurables ou repérables en variables ordinales par codage fait perdre une partie de l'information contenue dans les données initiales, mais l'uniformisation et la rigueur objective que la Taxonomie numérique recherche avant tout, est à ce prix.

Lors de nos premières tentatives d'application à deux groupes de Poissons, les Citharininae (Daget, 1966) et les Nototheniidae (Hureau, 1967) la standardisation des matrices de valeurs codes et l'emploi des distances taxonomiques n'étaient donc pas légitimes. Nous aurions dû passer de la matrice des valeurs codes à une matrice de similitude inter-UTO en utilisant seulement des statistiques d'ordre c'est-à-dire ne faisant intervenir que le rang de classement des valeurs codes.

Comme plusieurs caractères d'un même UTO sont souvent codés de façon identique, l'utilisation du *coefficient de corrélation de rang de Kendall*, qui tient compte des valeurs ex-æquo, semble tout indiqué. Ce coefficient, comme celui de Bravais-Pearson, est positif ou négatif et compris entre — 1 et + 1. Il existe des tables donnant les seuils de signification à 95 %.

Pour calculer le coefficient de corrélation de rang de Kendall entre deux UTO, on commence par remplacer les valeurs codes ex-æquo de chaque UTO (colonnes 2 et 3 du tableau I) par la moyenne de leurs rangs de classement. Ainsi pour le premier UTO il y a six valeurs codes égales à zéro auxquelles est attribué le rang de classement $3{,}5 = (1 + 2 + 3 + 4 + 5 + 6) / 6$, une valeur code égale à 1 à laquelle est attribuée le rang de classement 7, trois valeurs codes égales à 2 auxquelles est attribué le rang de classement $9 = (8 + 9 + 10) / 3$, etc. On classe ensuite les valeurs obtenues par ordre croissant pour l'un des UTO (colonne 2 du tableau II). On considère ensuite successivement chacune des valeurs de l'autre UTO (colonne 3 du tableau II) et on lui attribue autant de points positifs qu'il existe de valeurs supérieures après elle (colonne 4 du tableau II) et autant de points négatifs qu'il existe de valeurs inférieures après elle (colonne 5 du tableau II). Dans ce pointage, les valeurs égales à la valeur considérée ne sont pas comptées. De plus, si la valeur considérée correspond à un lot d'ex-æquo du premier UTO, le pointage ignore toutes les valeurs correspondant à ces ex-æquo.

Tableau I.

Données initiales		
caractères	UTO-1	UTO-2
1	0	0
2	0	0
3	2	3
4	4	3
5	5	6
6	0	1
7	0	1
8	1	1
9	2	5
10	2	4
11	0	3
12	0	1

Tableau II.

Données transformées			Pointage	
Caractères	UTO-1	UTO-2	+	—
1	3,5	1,5	6	0
2	3,5	1,5	6	0
6	3,5	4,5	5	0
7	3,5	4,5	5	0
12	3,5	4,5	5	0
11	3,5	8	3	1
8	7	4,5	5	0
3	9	8	1	0
10	9	10	1	1
9	9	11	1	1
4	11	8	1	0
5	12	12	0	0
			+ 39	— 3

Soit S_{ij} la somme algébrique des points. Dans l'exemple ci-dessus elle est égale à 39 — 3 = 36. Il est bien évident que l'on obtiendrait le même nombre en intervertissant les UTO, c'est-à-dire en classant les valeurs du second par ordre croissant et en faisant le pointage à partir des valeurs correspondantes du premier. Le coefficient de corrélation de rang de Kendall est donné par la formule :

$$t_{ij} = \frac{2\ S_{ij}}{\sqrt{n\ (n - 1) - \Sigma\ q_i\ (q_i - 1)}\ \sqrt{n\ (n - 1) - \Sigma\ q_j\ (q_j - 1)}}$$

q_i et q_j étant les nombres de valeurs ex æquo pour chacun des UTO. Dans l'exemple ci-dessus, n = 12 et par conséquent n (n — 1) = 132. Pour le premier UTO, il y a six valeurs codes égales à zéro et trois égales à deux ; on a donc $\Sigma q_i\ (q_i - 1) = 6.5 + 3.2 = 36$.

De même $\Sigma q_j\ (q_j - 1) = 4.3 + 3.2 + 2.1 = 20$

$$\text{donc } t = \frac{2.36}{\sqrt{132 - 36}\ \sqrt{132 - 20}} = \frac{72}{\sqrt{96.112}} = +\ 0{,}694.$$

En appliquant cette méthode et en partant de la matrice des valeurs codes relatives aux Citharininae (Daget, 1966, p. 380), on a obtenu la matrice d'intercorrélation suivante : (Matrice A)

Matrice A.

	UTO-1	UTO-2	UTO-3	UTO-4	UTO-5	UTO-6	UTO-7	UTO-8
UTO-1	+ 1,000	+ 0,451	— 0,307	— 0,109	— 0,198	— 0,018	+ 0,018	— 0,038
UTO-2		+ 1,000	— 0,082	+ 0,169	+ 0,237	+ 0,317	— 0,122	— 0,253
UTO-3			+ 1,000	+ 0,612	+ 0,732	+ 0,478	+ 0,637	+ 0,786
UTO-4				+ 1,000	+ 0,746	+ 0,867	+ 0,779	+ 0,569
UTO-5					+ 1,000	+ 0,694	+ 0,559	+ 0,540
UTO-6						+ 1,000	+ 0,607	+ 0,408
UTO-7							+ 1,000	+ 0,797
UTO-8								+ 1,000

L'analyse factorielle de cette matrice a été faite suivant la méthode de Hotelling (programme BMD 03M) sur ordinateur CDC 3600. Les résultats suivants ont été obtenus (Tableau III).

Le facteur général extrait 53,8 % de la variance totale. Les plus fortes saturations (0,800 à 0,907) affectent les UTO 3 à 8 qui constituent le genre *Citharinus* auct. moins *C. distichodoides*. Les saturations des UTO 1 et 2, respectivement — 0,162 et 0,048, sont nettement plus faibles. Ceci confirme nos conclusions précédentes à savoir que *C. distichodoides* ne doit pas être inclus dans le même genre ou sous-genre que les autres *Citharinus* car il est beaucoup plus proche de *Citharidium ansorgii* avec lequel il s'hybride parfois dans la nature (Daget, 1963).

Tableau III.

Facteurs	général	bipolaires						
UTO-1	— 0,162	0,685	0,684	0,064	— 0,141	0,054	— 0,093	— 0,017
UTO-2	0,048	0,912	— 0,208	0,260	0,208	— 0,070	0,079	0,007
UTO-3	0,842	— 0,260	— 0,035	0,402	0,120	0,113	— 0,178	0,029
UTO-4	0,907	0,186	— 0,105	— 0,282	0,048	— 0,063	— 0,075	— 0,201
UTO-5	0,850	0,136	— 0,296	0,242	— 0,308	— 0,131	0,027	0,020
UTO-6	0,801	0,373	— 0,202	— 0,318	— 0,035	0,255	0,034	0,098
UTO-7	0,858	— 0,075	0,365	— 0,215	0,102	— 0,222	— 0,013	0,136
UTO-8	0,800	— 0,291	0,432	0,186	0,031	0,090	0,200	— 0,072
Racines caractéristiques ...	4,301	1,652	0,972	0,554	0,187	0,164	0,094	0,094
Pourcentage de variance extrait	0,538	0,207	0,121	0,069	0,023	0,021	0,012	0,009

Le premier facteur bipolaire extrait 20,7 % de la variance totale. Les rangs de saturation classent les espèces du genre *Citharinus* suivant leurs moyennes vertébrales. Les saturations les plus faibles (— 0,291 et — 0,260) affectent les UTO 8 et 3 qui ont 40-42 vertèbres (mode 41) ; la saturation — 0,075 affecte l'UTO-7 qui a 41-43 vertèbres (mode 42) ; les saturations suivantes (0,136 et 0,186) affectent les UTO 4 et 5 qui ont 42-43 vertèbres ; enfin les saturations les plus élevées (0,373 et 0,912) affectent les UTO 6 et 2 qui ont 44-46 vertèbres (mode 45).

Dans l'espace factoriel correspondant au facteur général et au premier facteur bipolaire (74,5 % de la variance totale) les UTO 3 à 8 forment un groupe bien individualisé. Ceci justifie la reconnaissance de trois genres dans l'ensemble des Citharininae : *Citharinus* Cuvier 1817 (UTO 3 à 8), *Citharinops* Daget 1962 (UTO-2) et *Citharidium* Boulenger 1902 (UTO-1).

Le deuxième facteur bipolaire extrait 12,1 % de la variance totale. Les saturations positives (0,365 à 0,684) affectent les UTO 1, 7 et 8 dont la base de l'adipeuse a une longueur supérieure à 0,8 fois la distance entre l'adipeuse et la dorsale rayonnée. Les saturations négatives (— 0,035 à — 0,296) affectent les UTO 2 à 6 dont la base de l'adipeuse est inférieure à 0,8 fois la distance à la dorsale.

Le troisième facteur bipolaire extrait 6,9 % de la variance totale. Les saturations négatives affectent les UTO 4, 6 et 7 qui sont les espèces occidentales les plus évoluées du genre *Citharinus*, alors que les saturations positives affectent les UTO 1, 2, 3, 5 et 8 qui sont les espèces les moins évoluées du genre *Citharinus*, propres au bassin congolais et les espèces des genres *Citharidium* et *Citharinops* encore moins évoluées.

On n'a pas cherché à interpréter les autres facteurs bipolaires qui n'extraient que 6,5 % de la variance totale.

En résumé, l'analyse factorielle de la matrice d'intercorrélation obtenue à partir des coefficients de corrélation de rang de Kendall, calculés pour les huit

espèces de Citharininae prises deux à deux, et faisant intervenir douze caractères distinctifs, nous a permis de retrouver tous les résultats auxquels l'un de nous était arrivé précédemment par diverses méthodes d'approche (Daget, 1962, 1966). Dans le plan factoriel correspondant au facteur général et au premier facteur bipolaire (fig. 1) l'individualisation des trois genres *Citharidium*,

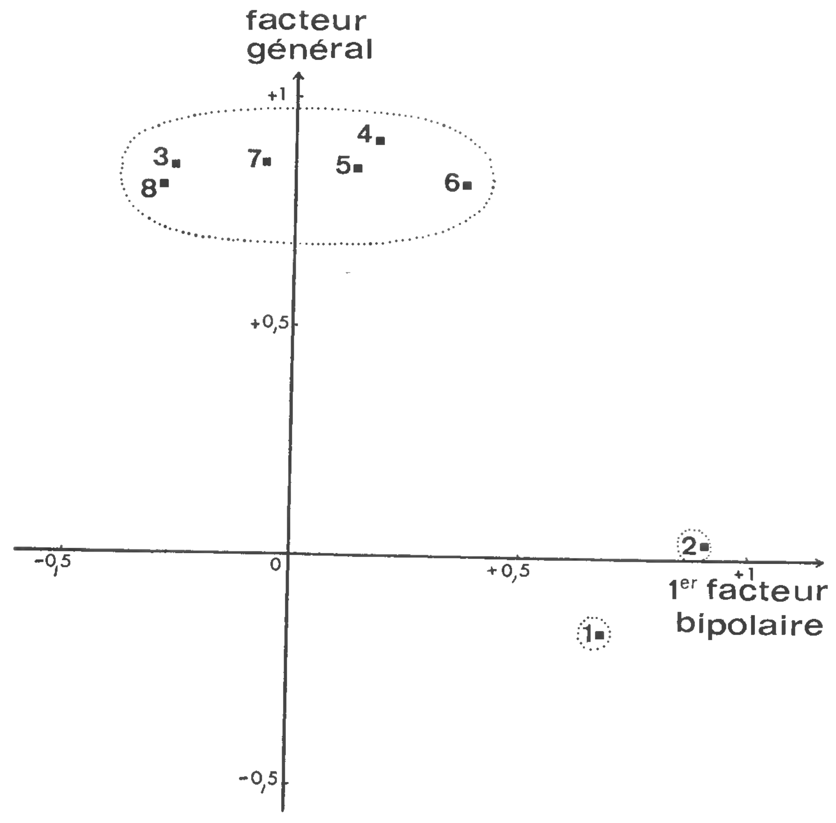


Fig. 1. — Position des points représentatifs des huit espèces de Citharininae dans le plan factoriel correspondant aux deux premiers facteurs principaux extraits. En abscisses, saturations du premier facteur bipolaire ; en ordonnées, saturations du facteur général.

1, *Citharidium ansorgii* ; 2, *Citharinops distichodoides* ; 3, *Citharinus macrolepis* ; 4, *Citharinus eburneensis* ; 5, *Citharinus congicus* ; 6, *Citharinus citharus* ; 7, *Citharinus latus* ; 8, *Citharinus gibbosus*.

Citharinops et *Citharinus* est nettement confirmée. Le deuxième facteur bipolaire rapproche *Citharinus latus* et *Citharinus gibbosus* de *Citharidium ansorgii* mais *C. latus* et *C. gibbosus* ne sont pas suffisamment isolés des autres *Citharinus* pour que le maintien du sous-genre *Citharinoides* (sensu Daget, 1962) soit justifié. Enfin le troisième facteur bipolaire regroupe les espèces suivant leur degré d'évolution et sépare, dans le genre *Citharinus*, les formes occidentales des formes congolaises.

Partant de la matrice des valeurs codes relative aux Nototheniidae (Hureau, 1967, p. 492), on a obtenu en appliquant la même méthode la matrice d'intercorrélation B.

Matrice B.

	UTO-1	UTO-2	UTO-3	UTO-4	UTO-5	UTO-6	UTO-7	UTO-8	UTO-9	UTO-10	UTO-11	UTO-12	UTO-13
UTO-1	+ 1,000	+ 0,688	+ 0,503	+ 0,384	+ 0,628	+ 0,302	+ 0,453	+ 0,206	— 0,574	+ 0,101	+ 0,192	+ 0,526	+ 0,544
UTO-2		+ 1,000	+ 0,656	+ 0,492	+ 0,722	+ 0,501	+ 0,450	+ 0,551	— 0,147	— 0,050	+ 0,380	+ 0,589	+ 0,587
UTO-3			+ 1,000	+ 0,437	+ 0,323	+ 0,213	+ 0,630	+ 0,356	— 0,072	+ 0,197	+ 0,406	+ 0,679	+ 0,390
UTO-4				+ 1,000	+ 0,328	+ 0,133	+ 0,381	+ 0,567	+ 0,147	— 0,083	+ 0,111	+ 0,454	+ 0,286
UTO-5					+ 1,000	+ 0,793	+ 0,340	+ 0,488	+ 0,228	— 0,379	— 0,098	+ 0,191	+ 0,361
UTO-6						+ 1,000	+ 0,182	+ 0,414	— 0,019	— 0,438	— 0,167	— 0,053	+ 0,066
UTO-7							+ 1,000	+ 0,292	— 0,200	+ 0,218	+ 0,277	+ 0,514	+ 0,467
UTO-8								+ 1,000	+ 0,273	— 0,497	+ 0,220	+ 0,334	+ 0,204
UTO-9									+ 1,000	— 0,212	— 0,184	— 0,155	— 0,220
UTO-10										+ 1,000	+ 0,634	+ 0,336	+ 0,233
UTO-11											+ 1,000	+ 0,707	+ 0,555
UTO-12												+ 1,000	+ 0,640
UTO-13													+ 1,000

Tableau IV.

Facteurs	général	bipolaires									
UTO-1	+ 0,771	— 0,035	— 0,519	+ 0,152	— 0,129	+ 0,192	— 0,152	+ 0,038	— 0,219	— 0,020	+ 0,105
UTO-2	+ 0,896	+ 0,171	— 0,078	— 0,155	— 0,060	— 0,015	— 0,170	+ 0,044	+ 0,093	— 0,217	— 0,224
UTO-3	+ 0,773	— 0,135	+ 0,140	+ 0,196	+ 0,320	— 0,237	— 0,193	+ 0,296	+ 0,142	— 0,062	+ 0,138
UTO-4	+ 0,603	+ 0,182	+ 0,394	+ 0,473	— 0,181	+ 0,260	— 0,159	— 0,276	+ 0,146	+ 0,048	+ 0,007
UTO-5	+ 0,648	+ 0,637	— 0,183	— 0,242	+ 0,194	+ 0,242	— 0,030	— 0,003	— 0,153	+ 0,009	+ 0,008
UTO-6	+ 0,406	+ 0,682	— 0,323	— 0,248	+ 0,195	— 0,240	— 0,066	— 0,223	+ 0,158	+ 0,186	+ 0,033
UTO-7	+ 0,688	— 0,147	— 0,017	+ 0,323	+ 0,411	— 0,103	+ 0,443	— 0,114	— 0,086	— 0,022	+ 0,075
UTO-8	+ 0,572	+ 0,496	+ 0,446	— 0,001	— 0,299	— 0,287	+ 0,117	— 0,083	— 0,151	— 0,092	+ 0,082
UTO-9	— 0,186	+ 0,418	+ 0,757	— 0,227	+ 0,317	+ 0,273	— 0,036	+ 0,077	— 0,061	— 0,003	+ 0,014
UTO-10	+ 0,088	— 0,868	+ 0,012	— 0,109	+ 0,319	+ 0,121	— 0,200	— 0,263	— 0,045	— 0,046	+ 0,024
UTO-11	+ 0,503	— 0,641	+ 0,270	— 0,401	— 0,134	— 0,229	— 0,038	— 0,162	— 0,055	— 0,016	+ 0,039
UTO-12	+ 0,781	— 0,412	+ 0,231	— 0,009	— 0,109	— 0,013	— 0,048	+ 0,198	— 0,083	+ 0,305	— 0,114
UTO-13	+ 0,712	— 0,290	— 0,064	— 0,285	— 0,165	+ 0,326	+ 0,361	+ 0,071	+ 0,215	— 0,019	+ 0,091
Racines caractéristiques	5,157	2,813	1,492	0,838	0,749	0,619	0,506	0,383	0,239	0,193	0,117
Pourcentage de variance extrait	0,397	0,216	0,115	0,064	0,058	0,048	0,039	0,029	0,018	0,015	0,009

Les résultats de l'analyse factorielle de cette matrice sont indiqués dans le tableau IV.

Le facteur général extrait 39,7 % de la variance totale. Les plus fortes saturations (0,688 à 0,896) affectent les UTO 1, 2, 3, 7, 12 et 13 c'est-à-dire les espèces : *Trematomus bernacchii*, *T. hansoni*, *T. loennbergii*, *Notothenia cyanobrancha*, *N. brevipectoralis* et *N. squamifrons*. Les plus faibles saturations (— 0,185 à 0,648) affectent les UTO 4, 5, 6, 8, 9, 10 et 11, c'est-à-dire les espèces : *Trematomus newnesi*, *Notothenia coriiceps neglecta*, *N. coriiceps coriiceps*, *N. rossii*, *N. macrocephala*, *N. acuta* et *N. gibberifrons*.

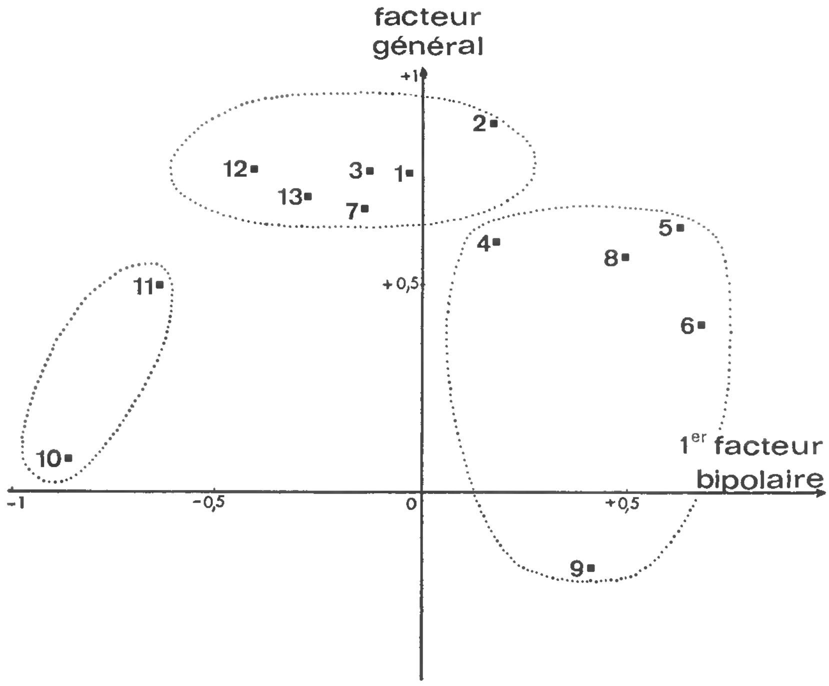


Fig. 2. — Position des points représentatifs de treize espèces ou sous-espèces de Nototheniidae dans le plan factoriel correspondant aux deux premiers facteurs principaux extraits. En abscisses, saturations du premier facteur bipolaire ; en ordonnées, saturations du facteur général.

1, *Trematomus bernacchii* ; 2, *Trematomus hansoni* ; 3, *Trematomus loennbergii* ; 4, *Trematomus newnesi* ; 5, *Notothenia coriiceps neglecta* ; 6, *Notothenia coriiceps coriiceps* ; 7, *Notothenia cyanobrancha* ; 8, *Notothenia rossii* ; 9, *Notothenia macrocephala* ; 10, *Notothenia acuta* ; 11, *Notothenia gibberifrons* ; 12, *Notothenia brevipectoralis* ; 13, *Notothenia squamifrons*.

Le premier facteur bipolaire extrait 21,6 % de la variance totale. Les saturations les plus élevées (0,682 à 0,182) affectent les UTO 4, 5, 6, 8 et 9 qui ont un espace interorbitaire large, compris 2,5 à 5 fois seulement dans la longueur de la tête. Les saturations moyennes (0,171 à — 0,412) affectent les UTO 1, 2, 3, 7, 12 et 13 dont l'espace interorbitaire plus étroit est compris 5 à 12 fois dans la longueur de la tête. Enfin les saturations les plus faibles affectent les UTO 10 et 11 dont l'espace interorbitaire très étroit est compris 12 à 16 fois dans la longueur de la tête.

Dans l'espace factoriel correspondant au facteur général et au premier facteur bipolaire (61,3 % de la variance totale) les UTO 1, 2, 3, 7, 12 et 13 forment

un premier groupe d'espèces ayant des caractères biologiques communs (fig. 2). On remarquera que l'UTO-7, *Notothenia cyanobrancha* se trouve très près des UTO 1, 2 et 3 qui correspondent aux *Trematomus* moins l'espèce *T. newnesi*, ce qui n'apparaissait pas sur les diagrammes publiés antérieurement par l'un de nous (Hureau, 1967). Ce rapprochement entre *N. cyanobrancha* et les *Trematomus* avait déjà été signalé dans un travail précédent (Hureau, 1966). Les UTO 10 et 11, *Notothenia acuta* et *N. gibberifrons*, forment un second groupe de deux espèces très voisines, à espace interorbitaire étroit et écailleux, mais géographiquement éloignées l'une de l'autre. Les UTO 4, 5, 6, 8 et 9 forment un troisième groupe dans lequel *Trematomus newnesi* se trouve rapproché de *Notothenia coriiceps coriiceps*, *N. c. neglecta*, *N. rossii* et *N. macrocephala*. Toutes ces formes sont voisines au point de vue morphologique (espace interorbitaire large et nu) et au point de vue biologique (migrations, œufs et alevins pélagiques).

En résumé, l'analyse factorielle de la matrice d'intercorrélation obtenue à partir des coefficients de corrélation de rang de Kendall, calculés pour treize espèces et sous-espèces de Nototheniidae prises deux à deux, et faisant intervenir treize caractères distinctifs, nous a donné des résultats en accord avec ceux déjà tirés d'études comparatives antérieures morphologiques et biologiques. Ils conduisent à conclure que les genres actuels *Notothenia* (espèce-type : *N. coriiceps*) et *Trematomus* (espèce-type *T. newnesi*) sont artificiels et que la répartition de l'ensemble des espèces au sein de la famille des Nototheniidae en genres et sous-genres devrait être entièrement refaite : le premier groupe défini ci-dessus pourrait constituer un premier genre, les second et troisième groupes pourraient former deux sous-genres d'un deuxième genre.

Résumé

Les auteurs ont calculé les coefficients de corrélation de rang de Kendall entre UTO pour obtenir une matrice d'intercorrélation. Ils ont ensuite effectué l'analyse factorielle suivant la méthode de Hotelling et représenté les UTO dans l'espace factoriel à deux dimensions correspondant au facteur général et au premier facteur bipolaire. Les résultats obtenus pour les Citharininae et les Nototheniidae sont en accord avec l'ensemble des observations déjà faites et montrent l'intérêt de cette méthode en Taxonomie numérique.

Summary

The authors have computed Kendall's rank correlation coefficients between OTUs to obtain an intercorrelation matrix. Then they have used the factor analysis according to the Hotelling's method and marked the OTUs in a factorial bi-dimensional space corresponding to the general and the first bipolar factor. Results obtained for Citharinines and Nototheniids check the whole of observations already made and point out the value of that method for Numerical Taxonomy.

Laboratoire de Zoologie
(Reptiles et Poissons) du Muséum.

RÉFÉRENCES BIBLIOGRAPHIQUES

DAGET J., 1962. — Le genre *Citharinus* (Poissons, Characiformes). *Rev. Zool. Bot. Afr.*, **66** (1-2) : 81-106, 12 fig.

DAGET J., 1963. — Sur plusieurs cas probables d'hybridation naturelle entre *Citharidium ansorgii* et *Citharinus distichodoides*. *Mém. I.F.A.N.*, Dakar, **68** : 81-83, 1 fig.

DAGET J., 1966. — Taxonomie numérique des Citharininae (Poissons, Characiformes). *Bull. Mus. Hist. nat.*, **38** (4) : 376-386, 2 fig.

HUREAU J. C., 1966. — Biologie comparée de quelques poissons antarctiques (Nototheniidae). *Bull. Inst. Océanog. Monaco*, 200 pp., 89 fig. (sous presse).

HUREAU J. C., 1967. — Taxonomie numérique des Nototheniidae (Poissons, Perciformes). *Bull. Mus. Hist. nat.* Paris, **39** (3) : 488-500, 2 fig.